ELSEVIER

Pattern Recognition Letters
journal homepage: www.elsevier.com

Automatic classification of flying bird species using computer vision techniques

John Atanbori[a,**], Wenting Duan[a], John Murray[a], Kofi Appiah[b], Patrick Dickinson[a]

[a]*University of Lincoln, School of Computer Science, Lincoln, LN6 7TS, UK*
[b]*Nottingham Trent University, School of Science and Technology, Nottingham, NG1 4BU, UK*

ABSTRACT

Bird populations are identified as important biodiversity indicators, so collecting reliable population data is important to ecologists and scientists. However, existing manual monitoring methods are labour-intensive, time-consuming, and potentially error prone. The aim of our work is to develop a reliable automated system, capable of classifying the species of individual birds, during flight, using video data. This is challenging, but appropriate for use in the field, since there is often a requirement to identify in flight, rather than while stationary. We present our work, which uses a new and rich set of appearance features for classification from video. We also introduce motion features including curvature and wing beat frequency. Combined with Normal Bayes classifier and a Support Vector Machine classifier, we present experimental evaluations of our appearance and motion features across a data set comprising 7 species. Using our appearance feature set alone we achieved a classification rate of 92% and 89% (using Normal Bayes and SVM classifiers respectively) which significantly outperforms a recent comparable state-of-the-art system. Using motion features alone we achieved a lower-classification rate, but motivate our on-going work which we seeks to combine these appearance and motion feature to achieve even more robust classification.

1. Introduction

Bird species are recognised as useful biodiversity indicators (Gregory, 2006; Harrison et al., 2014; Buckland et al., 2012). They are reponsive to changes in sensitive ecosystems, whilst populations-level changes in behaviour are both visible and quantifiable. Data about bird populations is therefore an important tool for ecologists in a wide range of environments and contexts, including farmland use, marine settings, and migration behaviour (Hammers et al., 2014; Johnston et al., 2014; Goodenough et al., 2014).

Current monitoring systems use manual methods of counting, or other more detailed observations which require trained personel to be deployed in sometimes quite inaccessible or hostile locations. This places practical limits on the quality and quantity of population-level data which can be collected. The objective of our work is to develop robust and reliable methods of collecting such data automatically, using computer-vision techniques. The successful development and deployment of such a system could enable collection of data on a scale not currently possible using manual methods, potentially allowing ecologists to conduct new types of scientific studies and investigations.

The work we present here focusses on the classification of species, using video data of individual birds in flight. There is some existing work which uses image analysis for species identification, but almost all (eg Marini et al. (2013)) use high-detail individual images for classification. This is less useful in the field, where automated systems are more likely to be deployed to monitor flying birds, which will inevitably present with poorer image quality (due to motion and distance). Flight patterns are known to vary with different species of birds (Briggs et al., 2012), so video data of flying bird also presents the opportunity to use motion features, and our longer-term objective is to combine both appearance and motion features for robust species classification.

In this paper we present our work to date, in which we have used two separate but extended sets of features (appearance and motion), with standard classifiers (Normal Bayes and Support Vector Machine). On our data set, which comprises videos of 7

**Corresponding author:
e-mail: `jatanbori@lincoln.ac.uk` (John Atanbori)

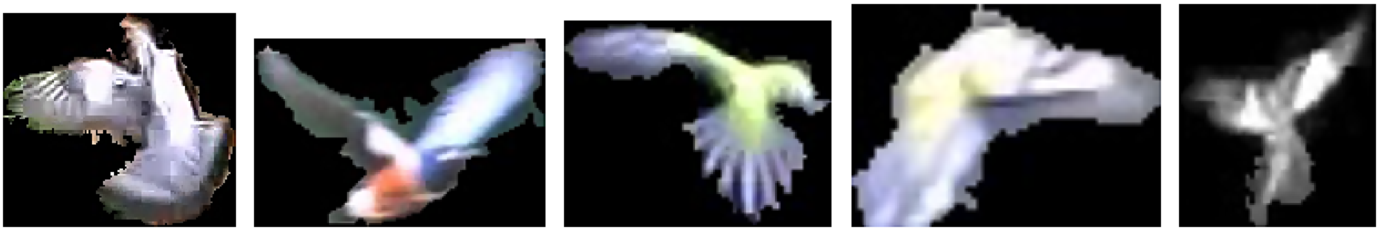

Fig. 1. Segmented birds from our flying birds data set using the method in (Zivkovic and van der Heijden, 2006). From left to right: Wood Pigeon, Superb Starling, Nanday Parakeet, Green Budgie and Cockatiel.

species in-flight, our proposed appearance classifier compares very favourably with exiting state-of-the-art image-based classifiers, and our motion classifiers provide strong encouragement that combined features will provide even more effective automated identification. The contributions of this paper are thus:

- We approach the challenging and unaddressed problem of in-flight species identification, with a proposed framework and data set.
- We present and evaluate a new set of appearance features, which we show experimentally is more reliable for classifying birds in flight than existing single-image based classifiers.
- We proposed a set of motion features we show to be effective, and which provide the basis for current ongoing work (to combine appearance and motion features).

The remainder of our paper is organized into the following sections. In Section 2 we review existing work on automated bird species classification, followed by an overview of our processing method in Section 3. We proceed in Section 4 to describe our feature extraction methods in detail (both appearance and motion), and conclude with experimental work, results, and discussion in sections 5 and 6 respectively.

2. Existing Work

A number of existing attempts to automate the identification of birds have used audio rather than visual signals, such as (Briggs et al., 2009; Neal et al., 2011; Lopes et al., 2011; Bardeli et al., 2010). The use of audio signals has some attractive features; species typically have distinctive calls, and no line of sight is necessary to detect audio. However, there is also significant disadvantages. Audio signals are sparse (an individual may emit no audio at all for extended periods), and it is not realistic to differentiate individuals in this way (e.g. for counting).

For this reason, a small but growing number of studies have looked at computer vision and image-based techniques which can potentially provide richer and more informative data (continuous position, behaviour, and other physical features). Works which use individual image-based identification include (Marini et al., 2013; Duberstein et al., 2012; Wah et al., 2011a; Duan et al., 2012; Berg and Belhumeur, 2013; Huang et al., 2013; Branson et al., 2014; Wah et al., 2011b). For our later evaluation we use the method proposed by Marini et al. (2013), which uses colour features for species identification, which was shown to be effective with a small number of classes. However, recognition rates dropped significantly with increasing numbers of classes, and this represents an ongoing challenge for robust identification.

Work by Huang et al. (2013) used a graphical model with saliency to classify 9 species of birds, by extracting Scale-invariant feature transform (SIFT) and colour features, which were train using different SVM classifiers and achieving 73.8% classification rate. SIFT and colour features (Wah et al., 2011b) works well in classification of bird species but again, is only tested for a small number of classes. The work presented in Welinder et al. (2010), classified bird species using size and colour histogram with bin size of 10 but the classification rate was also low, which was attributed to the fine-grained nature of the dataset.

Other works, such as Wah et al. (2011b) used a parts-based models and attribute-based learning. In this case, birds parts were annotated with the aid of human intervention, which is inappropriate for a fully-automated system for use in the field. Other examples include Duan et al. (2012) which discovered attributes automatically, but used human interaction to provide semantic names for those discovered attributes. These methods both provided good classification, but do require some degree of human intervention to identify parts: a tedious process that is also difficult to generalize to new species. In Berg and Belhumeur (2013); Yao et al. (2012) automated techniques were used for fine grain categorization of birds species. A method called Part-based one-vs-one features (POOF) was proposed in Berg and Belhumeur (2013), where the selection of regions for feature extraction was fully automatic, thus eliminating the human intervention. Yet this method still requires predefined birds parts which will be used for classification. The work in Branson et al. (2014), extracted object pose for fine-grained visual categorization to compute local image features, which were used for classification, achieving 75% correct classification rate.

These part-based methods have shown some success, but in all cases require some manual input, and also well-defined images in which the various parts present in a well-defined and identifiable way. These requirements are less appropriate for flying birds which may have less well defined object shapes, or in which specific parts may be obscured from view. A more recent work that was based on fine grain classification is the work in Berg et al. (2014).This work resulted in an online application called Birdsnap, which classifies various US bird species. Again, this was based on the improved version of the dataset in Wah et al. (2011b). Birdsnap uses a set of one-vs-most SVMs, based on POOF (Berg and Belhumeur, 2013). The one-vs-most

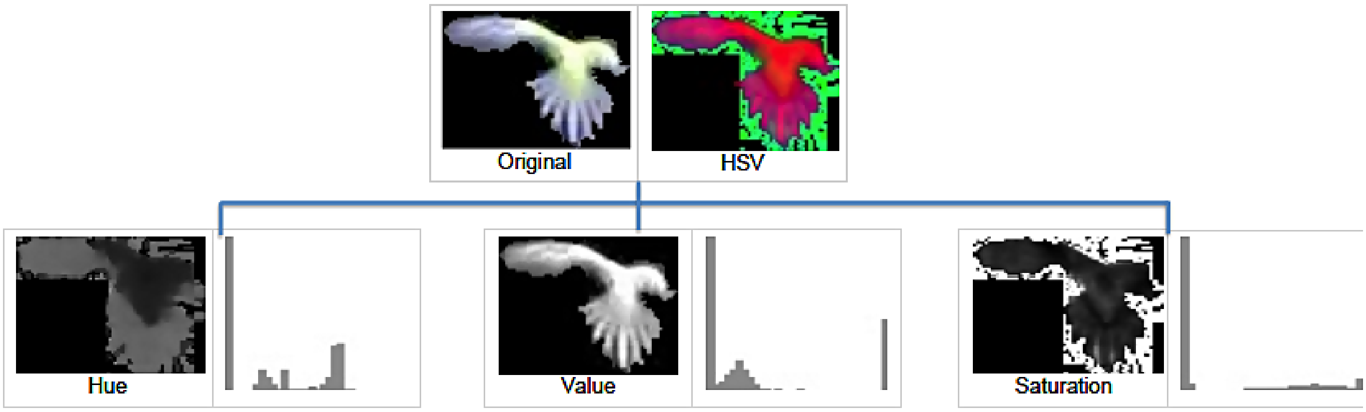

Fig. 2. Colour histogram from hue, saturation and value, which we extracted statistical featutres and concatenated with first order histogram probability to form our colour feature set.

SVM helped boost the results of their classification as compared to the work in POOF, which used a one-vs-one classifier. This work however, also involved some manual interaction from users, who marked the tail and eye of the bird species prior to classification.

One further significant work that performed fine-grained classification of bird species is that by Gavves et al. (2013). This approach performed an unsupervised alignment of birds images, and fitted an ellipse, which is used to obtain the birds parts for classification. The method achieved a very good classification rate, however images were prone to segmentation errors as they used the grab cut algorithm (which still requires user interaction).

All of the above mentioned methods use single images and appearance-based models for classification; however, bird species also exhibit distinguishing behaviours (flying, moving, poses, etc) which could also be used to help robust automated identification. This is particularly relevant to the identification of birds in flight, especially at distance where appearance-based features such as colour tend to attenuate, whilst motion-features remain discernible.

Motion-features associated with flight have not yet been well-explored for automated visual identification. The only significant relevant study is that of Duberstein et al. (2012) which explores the wing beat frequencies and flight trajectories patterns of bird species and also bats. However, this is limited to the broad categoriation of flight patterns rather than robust and specific species identification. This work using descriptive statistics extracted from each flight path including the minimum, maximum, mean, standard deviation, and quartiles of the data distribution as well as the interquartile range; they were able to achieve a coarse clustering of species. Atanbori et al. (2013) also presents preliminary work on classification of the flight trajectories of bats using similar analysis of wing-beat frequencies.

Trajectory and motion-based classification has been used more widely, however. For example, for the identification of people, fish, and vehicles (Duberstein et al., 2012; Beyan and Fisher, 2013, 2012; Anjum and Cavallaro, 2008; Li et al., 2006). The work of Beyan and Fisher (2013) in particular was used to classify the swimming trajectories of fish as normal or abnormal using statistical features extracted from 10 groups of features, which were concatenated for classification. Li et al. (2006) also investigated spatial and angular trajectory representations, using multi-features to classify vehicle trajectories.

One of the premises of our work is that the classification of birds in-flight may be achieved using either appearance or motion features, and that the combination of both feature types may achieve an overall more robust performance. No existing work has thus far attempted to encompass both of these in a robust or analytical way.

3. Material and Methods

3.1. Overview of Processing Method

The dataset used for this research is a video sequences of flying birds from 7 different species, recorded using a Casio Exilim ZR100 recording at 240 frames per second. The videos were recorded over different days from three different sites and species consist of more than ten individuals, apart from Superb Starlings, which had three.

We extracted the birds' silhouette (Fig 1.) using the background Gaussian mixture model proposed by Zivkovic and van der Heijden (2006). To detect the connected components, the contours where obtained from the binary image using the contour algorithm proposed by Suzuki et al. (1985). An oriented bounding box was fitted to each silhouette, and a selection of metrics (height, width and hypotenuse, centroid, silhouette and contour points) were measured. For any bird j tracked throughout n frames, our trajectory model is defined as the centre of the fitted bounded box (i.e. the centroid), given by the equation:

$$T_j = \{(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)\} \tag{1}$$

where T represents the trajectory and x and y are the coordinates of the centroid. Because the trajectory is noisy, we smoothen it by applying a box filter with 1 x 3 kernel.

Once the silhouettes are segmented, we extracted colour moments, shape moments, grayscale histogram, gabor filter and log-polar features. These features were concatenated to form one feature vector for classification of bird species by colour, shape and texture. For the trajectories and fitted curves, we extracted curvature scale space (CSS), turn based, centroid distance, vicinity and curvature based on sine and cosine features. These features together with wing beat frequencies were concatenated to form one feature vector for classification of bird species by their trajectories (as detailed in the next section).

In most cases, features were represented as statistical features, which provide information on the location, variability and appearance of the distribution of data and also to ensure classification can be performed in real time. The statistical features computed include the mean, standard deviation, skewness, kurtosis, energy, entropy, maximum, minimum, local maxima, local minima and number of zero crossings (details in the next section).

4. Feature Extraction

4.1. Appearance Features

To classify bird species, colour, shape and texture are important features. We represent colour features by colour moments and colour log-polar; shape features by shape moments; and texture features by Gabor filters and grayscale histogram. In this section, we present details of these features and what was extracted to represent them.

4.1.1. Color Moments Features

Histogram features have been used widely to describe colour images by extracting the histogram from various colour channels. To reduce the feature vector dimension and to make the system run in real-time, statistical measurements can be used to describe the colour histogram (Sergyan, 2008; Huang et al., 2010). A popular way to identify birds from video is using colour. Thus we deployed colour moments and calculated statistical features for speed of training and prediction.

To take advantage of colour we transform the colour images from RGB to HSV space before constructing the colour histogram (Fig. 2.). The histogram was then built for each colour channel separately and the first-order histogram probability (Sergyan, 2008) computed. After that, five statistical features were obtained: Mean, Standard Deviation, Skewness, Energy and Entropy for each of the colour channels. We concatenated the first-order histogram probability, to form the colour moments. There were 35 features for hue, 37 each for saturation and value. In total 109 statistical features were used to represent the colour moments.

4.1.2. Shape Moments Features

One important feature for identifying birds is the shape of the bird species. To describe the shape of an object various image moments can be extracted from the image contours (Du et al., 2007). An image moment is a weighted average of the image pixels' intensities. We used two moments in the extraction of shape features for bird species classification, which include spatial (raw) moments and Hu moments (Fig. 4). Hu moments are invariant to some transformations, such as rotation, scaling, and translation (Martín et al., 2010) and are therefore well suited for flying bird species classification.

To represent shape information, we extracted seven features from Hu moments and ten from spatial (raw) moments. In total 17 statistical features were used to represent shape moments.

4.1.3. Grayscale Histogram Features

Distributions of gray levels in images are commonly used in image analysis and classification. This is often done by representing the gray level distribution as histogram. For example, statistical moments of the gray scale histogram are used as features for the classification of fish species in Spampinato et al. (2010). In our work, we used Grayscale histogram features as texture to complement Gabour fetaures, which gave information about the spatial arrangement of intensities in the bird species video. They can be used online and have been used by many content based retrieval systems as features for classification

For texture features we converted the segmented image into a grayscale image and used it to form a histogram with 256 bins. We then calculated statistical moments features similar to Spampinato et al. (2010) from the histogram, which were used to form grayscale histogram features. In total eight features including mean, standard deviation, skewness, kurtosis, energy, entropy, Hu's 2nd and 3rd moments, were extracted to represent grayscale histogram features.

4.1.4. Gabor Wavelet Features

Gabor wavelets have been applied to many feature extraction problems (Huang et al., 2010; Ou et al., 2010; Parvin et al., 2012; Spampinato et al., 2010) due to its salient visual properties such as spatial localization, frequency characteristics and orientation. Assume an image I given by I(x, y), the Gabor wavelet transform is the convolution between the function g and image I, given by equation.

$$g(x, y; \theta, \lambda, \psi, \gamma, \sigma) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \exp\left(i\left(2\pi\frac{x'}{\lambda} + \psi\right)\right) \quad (2)$$

where:

- $x' = x\cos\theta + y\sin\theta$
- $y' = -x\sin\theta + y\cos\theta$
- and θ , λ, ψ, γ and σ are orientation, wavelength, phase, aspect ration and standard deviation respectively.

Gabor filter is scale invariant, as the size of the convolution kernel does not affect the output image. Gabor filters provide us with information about the spatial arrangement of intensities in the bird species. They can be used online and have been used by many content based retrieval systems.

To extract Gabor wavelets features, we used four orientations with one scale, thus obtaining four processed images (as shown in Fig. 3.). For each processed image we compute five statistical features including the mean, standard deviation, skewness, energy and entropy. In total 20 statistical features were used to represent Gabor wavelet features.

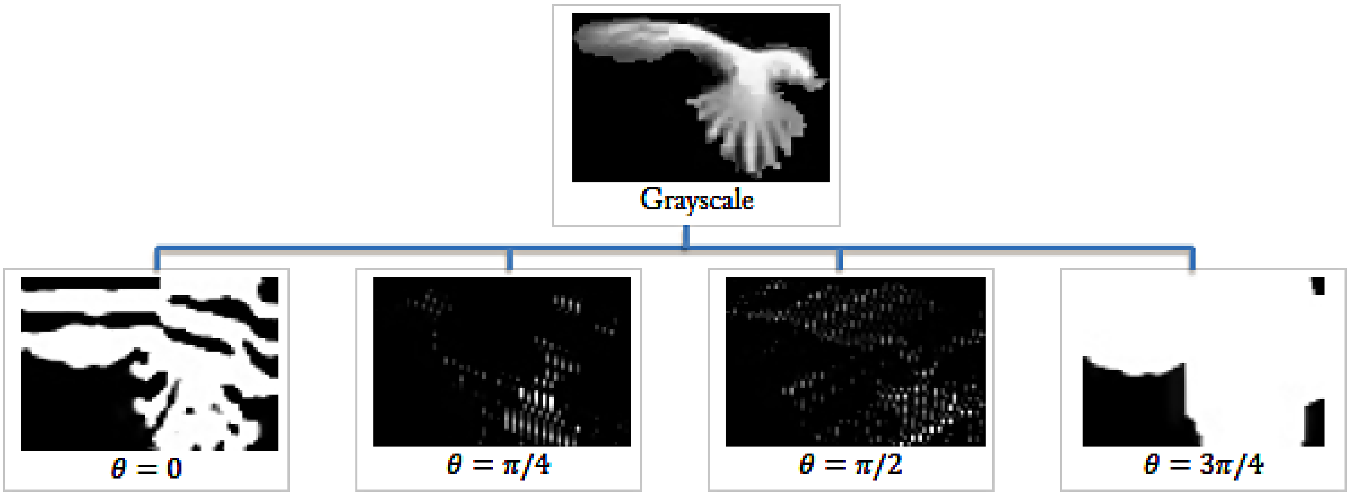

Fig. 3. Gabor filter features for four orientations. Five Statistical features were extracted from these and the results concatenated to form a feature vector.

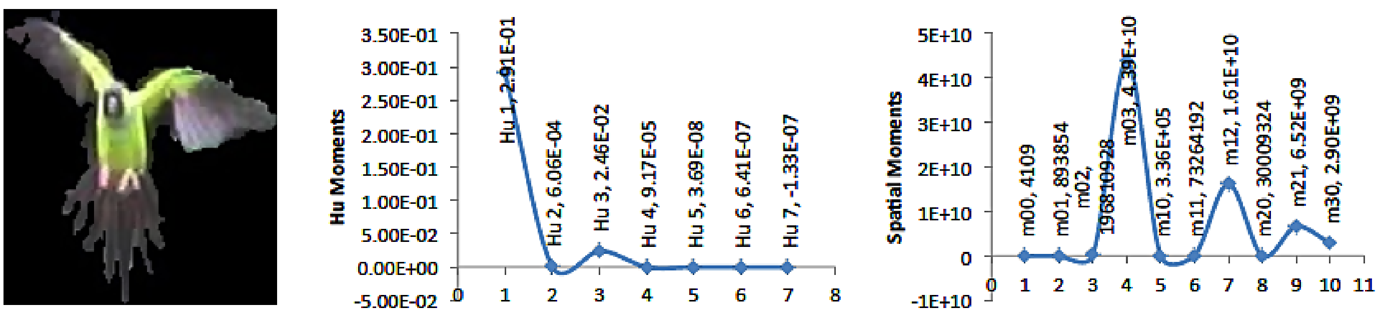

Fig. 4. Hu (in middle) and Spatial (at right) Moments plots of a segmented Nanday Parakeet's (at left)

4.1.5. Colour Log-Polar Features

Log-Polar transform can be used to eliminate effects of rotation and scale in input image, by converting the image into a log-polar image before processing (Pun and Lee, 2003). One common problem with flying birds video is the rate at which the angle of view and scale of the birds species changes rapidly in the scene. To alienate these effects, we used log-polar transform, by converting the image into a corresponding log-polar image.

Considering an image *src* with cartesian coordinate denoted by *src*(x, y) , this can be transformed into log-polar form in a destination image *dst* as *dst*(θ,ρ) given by

$$dst\,(\theta,\rho) \leftarrow src\,(x,y)\,for \begin{cases} \rho = \log\sqrt{x^2+y^2} \\ \theta = \arctan\left(\frac{x}{y}\right) if\, x > 0 \end{cases} \tag{3}$$

We extracted log-polar feature based on three channels hue, saturation and value. This was used to complement our colour moment features. The segmented image is converted into HSV colour space and applied log-polar transformation separately to each channel (see Fig. 5.). Five statistical moment features were acquired from each channel including mean, standard deviation, skewness, entropy and energy. Similar as previous process, these features were concatenated into a total of 15 features to form a colour log-polar features for our bird species classification. Our approach is different from existing approaches because we considered colour information whiles computing the log-polar features.

4.2. Trajectory Features

Trajectories of some bird species varies significantly and may therefore be used as features for classification of these species. In this section, we present details of the selected features used to represent bird species trajectory (motion).

4.2.1. Curvature Scale Space

Curvature scale space (CSS) is rotation and translation invariant and useful in distinguishing trajectories by their concave and convex shapes (Beyan and Fisher, 2013; Mai et al., 2010; Bashir et al., 2006). To have an affine representation of birds' trajectories in the form of their constituent sub-trajectories, we used the CSS features, which are robust representation of trajectory shape even in the presence of noise. The curvature at every point on the trajectory is calculated using the equation:

$$K_i = \frac{x_i' y_i'' - y_i' x_i''}{\left(x_i'^2 + y_i'^2\right)^{\frac{3}{2}}} \tag{4}$$

where x_i', x_i'', y_i' and y_i'' are first and second derivatives of x_i and y_i respectively

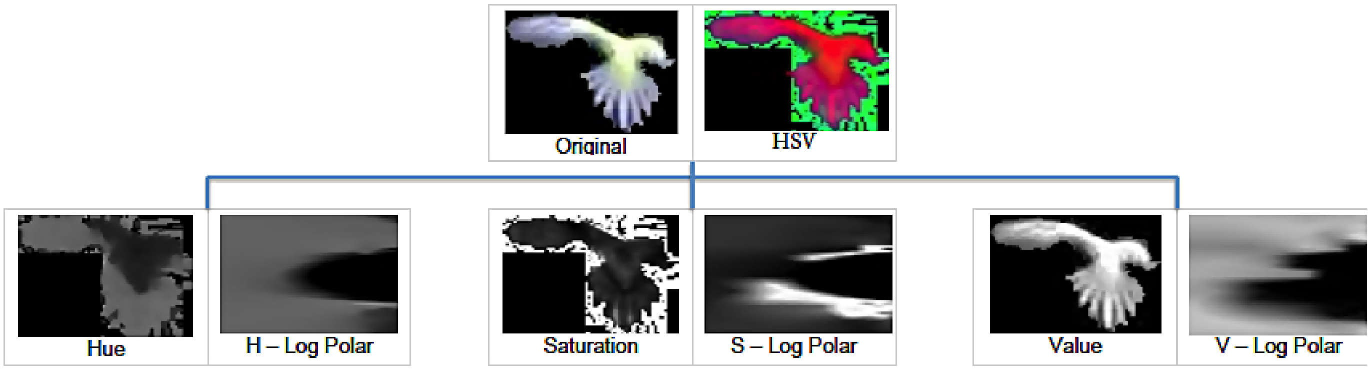

Fig. 5. Colour log-polar features obtained by extracting hue, saturation and value log-polars then extracting statistical features, which where concatenated to form a colour log-polar feature vector.

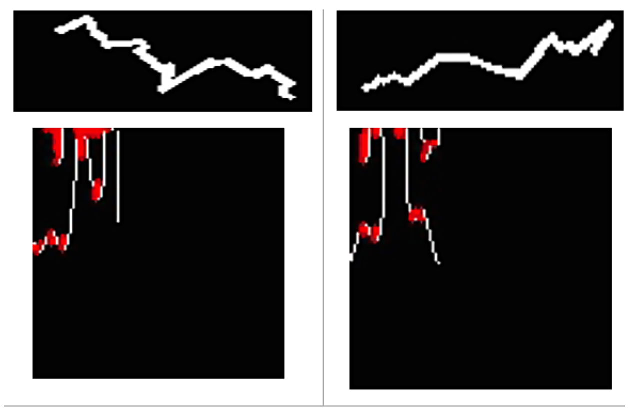

Fig. 6. Flight trajectories (above) of wood pigeon and itś corresponding CSS images (below).

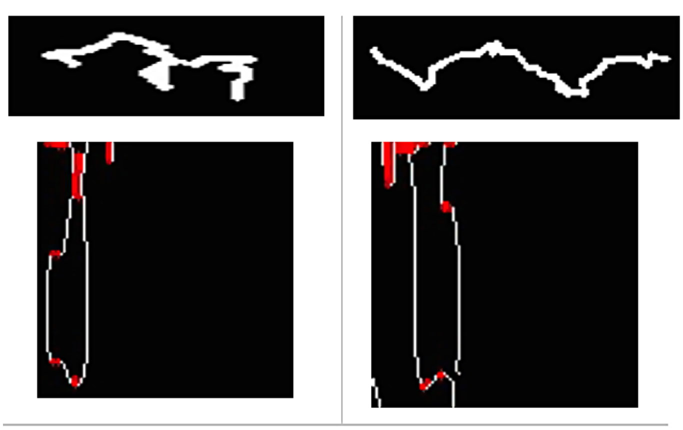

Fig. 7. Flight trajectories (above) of house martins and itś corresponding CSS images (below).

CSS is computed by applying a Gaussian smoothing kernel iteratively with different standard deviations. At each level of the standard deviation, a corresponding zero crossing from the second derivative of the trajectory are recorded to form the CSS image (Van De Sande et al., 2010). This process continues until there are no zero crossings and the trajectory becomes a convex curve.

The locations of CSS maxima in terms of their temporal ordering and the scale of concavity represent the trajectory. Each of the maxima location corresponds to a concavity in the shape of trajectory (Bashir et al., 2006).

We then extracted ten statistical features from the absolute curvature computed from equation (4), the number of curves in constructed CSS image (see Fig. 6 and 7), the total length of curves in CSS image and ten statistical moments from the CSS maxima. In total 22 features were used to represent CSS.

4.2.2. Turn Based Features

In order to obtain the shape of each bird flight trajectory, the trajectory turnings were calculated. This was obtained by calculating the slope of the bird trajectory between two consecutive frames as given in Li et al. (2006) for vehicle trajectory classification and in Beyan and Fisher (2013) for fish trajectory clustering. The calculation follows equation (5), where dx and dy are the change along the x and y axis respectively.

$$\theta(i) = \frac{dy(i)}{dx(i)} \quad if \quad dx(i) \neq 0 \tag{5}$$

Since we used 32 trajectory points, there were in total 30 features used to represent turn between trajectory points.

4.2.3. Wing Beat Frequency Based features

The periodic motion features associated with beating wings vary among species (Lazarevic et al., 2008), and may provide a useful discriminating feature for classification. This includes both short-scale features (frequency while flapping), but also others: for example some species characteristically mix flapping and gliding. In Atanbori et al. (2013), we showed that for bat species a bounding box fitted to the silhouette of a tracked individual can be used to accurately measure such periodicity features. We used this idea to extract the bird periodic motion as a 1D signal broken into short overlapping windows to cover the three metrics, height, width and diagonal of the bounding box. We compute the Fast Fourier Transform (FFT) (eqn. 6) for each of these time signals (*f(x)*) and extracted the first nine frequencies of the FFT excluding the DC component. The frequencies were then concatenated to form a feature vector of size 27 to represent the wing beat frequencies features of the birds.

$$F(u) = \frac{1}{N}\sum_{i=0}^{(N-1)} f(x)\, e^{-2\pi i x u/N} \tag{6}$$

where k is frequency, *f(x)* is the signal in the spatial domain and *F(u)* in the Frequency Domain (encoding both amplitude and phase).

4.2.4. Centroid Distance Function (CDF)

The centroid distance function (CDF) is an invariant representation of the shape of data (Beyan and Fisher, 2013; Bashir et al., 2006). Since the flying birds trajectory are subject to rotational deformation, CDF representation is a good way to represent this information. We calculated CDF by finding the centre point of the trajectory and calculating the distance of each trajectory from this centre point.

$$CDF_i = \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2} \;\; for\ i = 0, 1, ...N-1 \tag{7}$$

where: N is the total number of trajectory points.
$x_c = \frac{1}{N}\sum_{j=0}^{N-1} x_j$ and $y_c = \frac{1}{N}\sum_{j=0}^{N-1} y_j$

The computed CDF were normalized and their statistical moments used to represent CDF features. In total ten features including mean, maximum, minimum, standard deviation, number of zero crossings, number of local minima and maxima, skewness, energy and entropy of 2D CDF were extracted to represent the birds trajectory shape.

4.2.5. Vicinity

Vicinity was used in Liwicki et al. (2006) for handwriting recognition, and recently adopted for extracting features in Beyan and Fisher (2013) to classify abnormal trajectories of fishes by considering vicinity curliness, aspect, and linearity. Since they consist of features extracted from each point and take into consideration their neighbouring points and are very robust to noisy data, they were very suitable for the fly birds trajectories.

We computed vicinity features (Liwicki et al., 2006) for curliness, aspect, slope and linearity, which were normalized and used to calculate statistical moments. In total 40 vicinity features including mean, maximum, minimum, standard deviation, number of zero crossings, number of local minima and maxima, skewness, energy and entropy were extracted to represent the birds trajectory shape.

4.2.6. Curvature

Curvature features used in Li et al. (2006); Anjum and Cavallaro (2008) can help represent the shape of birds' trajectories. It is computed as the cosine of the angle between the lines to the previous and the next point. Assume a trajectory of points P_i for $i = 1$ to n, where n is the total number of points in the trajectory, then the cosine of angle between the lines $|P_{i-1}P_i|$ and $\|P_iP_{i+1}|$ is given by:

$$K_i = \cos^{-1}\left(\frac{\left(|P_{i-1}P_i|^2 + |P_{i-1}P_{i+1}|^2 + |P_iP_{i+1}|^2\right)}{(2\,|P_{i-1}P_i|\,|P_{i-1}P_{i+1}|)}\right) \tag{8}$$

Where: $|P_iP_{i+1}| = \sqrt{(P_ix - P_{i+1}x)^2 + (P_iy - P_{i+1}y)^2}$

The computed normalized curvature values were used to calculate statistical moments to represent curvature features. In total ten features including mean, maximum, minimum, standard deviation, number of zero crossings, number of local minima and maxima, skewness, energy and entropy of curvature were extracted to represent the bird's trajectory curves.

5. Experiments

We performed a series of experiments to evaluate the effectiveness of our proposed appearance and motion feature sets. We evaluated the appearance and motion sets independently, and compared with the results obtained using the colour features proposed by Marini et al. We tested all feature sets using both a Normal Bayes classifier, and the Support Vector Machine (SVM) used by Marini et al. (2013).

To facilitate evaluation, we used a simple cross-validation scheme based on a 70% training set, and 30% test set. For all experiments, we repeated the evaluation for four different test sets, and averaged to obtain the results presented in this section. For each experimental run we sampled individual image frames (from the training and test set) for which we extracted corresponding appearance and motion features. We used an average of 16,400 image frames from each training set, and an average of 7,221 from each test set. The entire dataset comprises 162 videos, each containing between 0.25 and 5 seconds of high speed video. We also investigated, experimentally, the effect of sample size on classification rates, using our appearance features set and the Normal Bayes classifier.

The Normal Bayes classifier assumed a Gaussian mixture model over the whole training data distribution, one component per class, and estimated parameters from the training data. Our SVM classifier is comparable to that used by Marini et al. (2013), and implemented using a radial basis function kernel, with the gamma and cost parameters optimized using a 5-fold grid search for parameterisation and validation.

Firstly, we set up two Normal Bayes classifiers, one for our concatenated appearance features and the other for the feature set used by Marini et al. (2013), for a direct comparison. We then repeated these experiments using the SVM classifier: as a complete grid search is time consuming, we applied the approach used in Hsu et al. (2003), by performing a coarse grid search first and after a good region on the grid was found, we perform a finer grid search with that region. Finally we used both the Normal Bayes and SVM classifiers with our motion feature set, for comparison with our proposed appearance features.

6. Results

6.1. Normal Bayes Classification Results

Tables 1 and 2 show the confusion matrices of results obtained using our appearance features and those of Marini et al. (2013) respectively, using the Normal Bayes classifier. The overall average classification rate for our method was 92%, which out-performed Marini et al. (2013) (68%) considerably.

Table 1. confusion matrix of classification using our appearance features on Normal Bayes Classifier (our method). These is the average of our four different testings

	House Martin	Wood Pigeon	Superb Starling	Nanday Parakeet	Cockatiels	Black Bird	Green Budgie	(%)
House Martin	1192	1	12	1	16	1	3	97
Wood Pigeon	3	1136	1	20	3	1	10	97
Superb Starling	2	0	579	0	29	0	2	95
Nanday Parakeet	9	26	12	947	25	65	21	86
Cockatiels	84	14	27	6	820	13	15	84
Black Bird	0	5	0	2	1	551	14	96
Green Budgie	14	2	3	16	19	29	706	90
Total (%)								92

Table 2. Confusion matrix of classification using colour features on Normal Bayes classifier (method in (Marini et al., 2013))

	House Martin	Wood Pigeon	Superb Starling	Nanday Parakeet	Cockatiels	Black Bird	Green Budgie	(%)
House Martin	1098	0	77	0	5	4	41	90
Wood Pigeon	30	714	22	1	2	3	402	61
Superb Starling	8	10	572	0	8	1	13	94
Nanday Parakeet	216	145	38	166	11	110	420	15
Cockatiels	234	3	231	0	379	10	123	39
Black Bird	17	0	1	0	1	491	63	86
Green Budgie	44	0	5	0	3	19	717	91
Total (%)								68

From the four testings we performed, the maximum classification rate was 95% and the minimum was 89% using our features while that of Marini et al. (2013) was 76% and 64% respectively. Based on the averages, using our features, the best classification was obtained for House Martins (97%) whiles the lowest was for Cockatiels (84%). Using Marini et al. (2013), based on the averages, the best classification rate was obtained for Superb Starling (94%) and the lowest for Nanday Parakeet, at only 15%. Nanday Parakeets and Green Budgies have very similar colour features (Fig. 1), and it appears that Nandy Parakeets are typically misclassified as Green Budgies when relying on colour alone. Also using the features from Marini et al. (2013), segmentation error can cause classification errors. In table 2, Nanday Parakeets (green colour) and some bird species which were filmed with green backgrounds (green grass and trees) had contaminated segmentation, especially wood pigeons (see Fig. 8.) and were therefore misclassified. Our method remains robust to these specific errors.

Our features worked well with both classifiers than the one used in Marini et al. (2013), however the data set performed better on the Normal Bayes Classifier than the SVM. Misclassification of bird species by observation was due to illumination and similar colour patterns in some species.

6.2. SVM Classification Results

Tables 3 and 4 shows the confusion matrices of the results using our method, and that in Marini et al. (2013), using the SVM classifier. The overall classification rate obtained using our method was 89%, out-performing Marini et al. (2013) again (71%). In our method the highest classification was for Green Budgies (94%) whilst the lowest was for Cockatiels (still good, at 85%). For comparison, using (Marini et al., 2013), the highest classification rate was for Black Birds (93%) and the lowest again for Nanday Parakeet (25%). Again, this demonstrates the weakness of using only colour-based features.

Table 3. Confusion matrix of classification using our appearance features on SVM classifier (Our Method)

	House Martin	Wood Pigeon	Superb Starling	Nanday Parakeet	Cockatiels	Black Bird	Green Budgie	(%)
House Martin	1302	3	59	4	131	2	2	87
Wood Pigeon	0	1203	0	73	9	0	28	92
Superb Starling	9	0	625	0	26	0	4	94
Nanday Parakeet	2	19	4	1023	16	73	63	85
Cockatiels	59	13	28	0	817	12	35	85
Black Bird	5	2	1	6	0	535	71	86
Green Budgie	2	1	5	29	6	3	811	95
Total (%)								89

Table 4. Confusion matrix of classification using colour features on SVM classifier (proposed in (Marini et al., 2013))

	House Martin	Wood Pigeon	Superb Starling	Nanday Parakeet	Cockatiels	Black Bird	Green Budgie	(%)
House Martin	1042	20	3	5	513	15	5	65
Wood Pigeon	8	1155	5	18	26	2	99	88
Superb Starling	9	27	584	3	31	1	9	88
Nanday Parakeet	8	654	57	297	135	7	42	25
Cockatiels	96	28	60	8	742	13	17	77
Black Bird	0	3	4	3	27	576	7	93
Green Budgie	12	119	7	25	71	80	543	63
Total (%)								71

6.3. Motion Features Classification Results

Tables 5 and 6 show the confusion matrices obtained using our motion feature set with both the Normal Bayes classifier and SVM respectively. The classification using motion features showed rather interesting results, with an overall correct classification rate of 37% based on the average of our four testing. The highest rate was again obtained for Green Budgies (66%), and the lowest for Black Birds (8%). We also observed that the maximum classification rate from the four testings was 41% and the minimum 33%. However, the overall classification rate (from all four testings) for the majority of species were above 40% using the Normal Bayes classifier, which evidences the ability of these features to provide additional differentiation.

6.4. Sample Size Per Class

We also observed that from a sample of 200 to 700 per class that the classification rate increases sharply. Beyond 700 samples the improvement is less significant, but peaks at approximately 2000 samples per class (Fig. 9.).

Table 5. Confusion matrix of classification using motion features with Normal Bayes Classifier

	House Martin	Wood Pigeon	Superb Starling	Nanday Parakeet	Cockatiels	Black Bird	Green Budgie	(%)
House Martin	347	61	60	160	200	34	138	35
Wood Pigeon	93	293	54	271	250	16	66	28
Superb Starling	28	2	185	17	21	1	140	47
Nanday Parakeet	107	121	70	320	220	14	67	35
Cockatiels	110	46	43	64	316	8	222	39
Black Bird	95	35	34	83	115	36	36	8
Green Budgie	29	3	61	26	43	3	314	66
Total (%)								37

Fig. 8. Segmented Wood Pigeons contaminated with green background.

Table 6. Confusion matrix of classification using motion features with SVM Classifier

	House Martin	Wood Pigeon	Superb Starling	Nanday Parakeet	Cockatiels	Black Bird	Green Budgie	(%)
House Martin	273	192	54	103	495	143	126	20
Wood Pigeon	39	426	63	146	428	39	48	36
Superb Starling	19	3	153	10	69	5	188	34
Nanday Parakeet	24	281	59	75	405	71	99	7
Cockatiels	47	87	18	21	422	32	182	52
Black Bird	79	77	13	54	182	71	20	14
Green Budgie	46	2	41	2	129	9	322	58
Total (%)								32

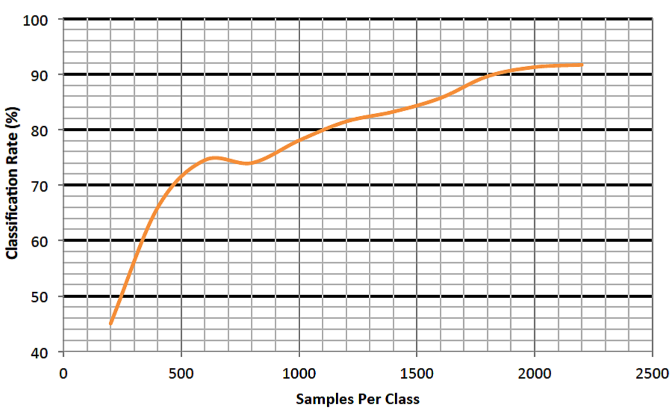

Fig. 9. Sample size per class against classification rate (%)

6.5. *Performance Evaluation*

In this section, we empirically compare the computational performance of our proposed classifiers with those of Marini et al. (2013), and determine whether they are capable of running in real-time on standard computational hardware. We performed these experiments on a Mac book pro laptop running OS X 10.9.5, with 2.5 GHz Processor and 4 GB Ram. The algorithms and classifiers were all written in C++ with XCode 5.1.1 and OpenCV 3.0. We tested both the classification and recognition phases, separately.

To compare performance for the training phase, we used a software timing function (millisecond accuracy) and recorded the time in seconds taken to build Bayes and SVM classifiers for our appearance feature set, motion-feature set, and for Marini et al.'s algorithm. The timings include both image feature extraction and training phases, using 70% (16,400 image frames) of our total data set (as according to our reported experimental setup). We ran each timings for the same classifier 50 times so that we could estimate a mean value, and also report the standard deviation, minimum, and maximum timings in each case. Results are presented in table 7.

Table 7. Recorded times for classier training.

	Appearance		Motion		Marini	
	Bayes	SVM	Bayes	SVM	Bayes	SVM
Mean	157.51	152.10	202.65	202.72	85.74	87.45
Min	141.95	136.67	186.38	186.35	66.07	67.66
Max	202.33	196.31	238.03	238.31	157.71	159.54
σ	11.00	10.82	14.94	15.08	19.48	19.64

Table 8. Classification times for 1,500 birds in seconds and the estimated classification times for a single bird in milliseconds.

	Appearance		Motion		Marini	
	Bayes	SVM	Bayes	SVM	Bayes	SVM
Mean/1500 Birds	13.43	13.43	18.01	18.07	7.67	7.72
Min/1500 Birds	12.05	12.05	16.56	16.61	5.91	5.96
Max/1500 Birds	17.37	17.36	21.16	21.25	14.05	14.16
σ/1500 Birds	0.96	0.96	1.33	1.34	1.73	1.75
Mean/Birds (in ms)	8.95	8.95	12.01	12.05	5.11	5.15

To compare performance for the classification phase, we performed a similar set of experiments. For each classifier we took a set of 1500 individual birds, and recorded the time taken to classify the entire set. This includes both the feature extraction, and the actual classification using Bayes and SVM and the appearance features, motion features, and Marini et al.'s feature set. Again, we repeated our experimental runs 50 times, and obtained an average time (for all 1500 birds), standard deviation, minimum and maximum values, as shown in table 8. We also divided the mean by 1500 to calculate an indicative time for a single bird which represents in the field performance for our system.

From table 7, it can be seen that using our appearance features, training took on average 157.51 seconds with the Bayes Normal classifier, and 152.10 with SVM. Marini et al. (2013) was faster with 85.74 seconds and 87.45 seconds. Our motion features took 202.65 seconds and 202.72 seconds respectively. As expected, the training times are slower with our larger feature sets, but very acceptable for off-line training. Comparison of the two classifiers (Bayes vs SVM) shows little variation using the same features, and this illustrates clearly that it is the feature extraction process, rather than the training process, which dominates the time required for the training phase. The training complexity of SVM is of the order $O(nd)$ Keerthi et al. (2006), and Normal Bayes training complexity has been shown similarly to be of order $O(nd)$ Zheng and Webb (2005), where n is the number of training samples and d is the feature dimension. Given our set of defined features, we expect the complexity of the training phase to be approximately linear in the number of training samples, since both the dominant feature extraction process, and training, are both linear.

From table 8, we observe that Marini et al. (2013) returns a classification slightly faster than either our appearance-based or motion-based classifiers (with either Bayes or SVM). The classification complexity of SVM and Normal Bayes are both also of the order $O(d)$ Zheng and Webb (2005); Keerthi et al. (2006), which means that the time it takes to classify a bird is linearly dependent on the feature dimension. However, our results indicate that the choice of either Bayes Normal or SVM makes relatively little difference to performance, which again is heavily dominated by the feature extraction process. Crucially, and despite an approximate factor of 2 increase in processing time when compared with Marini et al. (2013), both our motion and appearance-based feature sets are able to return a single classification in less than $10ms$. Even combining both appearance and motion feature sets, we expect a single bird classification process to take around $20ms$, which is more than suitable for real-time application.

7. Conclusion

In this paper we have described our proposed feature sets (appearance and motion) for automated species classification of flying birds, and presented supporting experimental results in which we compared our appearance features with those of Marini et al. (2013) which uses only colour-based features. The classification of flying birds is a particularly challenging context for automated species identification, and no existing work has yet addressed this problem directly.

We used both SVM and Normal Bayes classifiers to evaluate our feature sets experimentally, using our video data set (which covers 7 species of flying birds). Our results show that using both SVM and Normal Bayes classifiers, our proposed feature sets out performs the recent state-of-art colour feature classifier presented by Marini et al. (2013). Specifically, the overall correct classification using Normal Bayes was found to be 92% against 68%, and 89% vs 71% for SVM. Our results demonstrate that the utility of shape and colour features in this context, particularly for resolving ambiguities between species with similar colouration.

We have further considered the use of motion features for automated bird species classification: again, a research area with very little, if any, existing work. Using our proposed feature set, across 7 species, we achieved a classification rate of 37% using a Normal Bayes classifier, ranges from a 8% to 66% correct classification across species. The majority of species had a correct classification rate of greater than 40% in all our testings. We have consequently established the utility of such features, and postulate that they will be most effective at distance, where colour features are more likely to attenuate.

We have also evaluated the computational performance of our feature set against that of Marini et al. (2013). Eventhough Marini et al. (2013) returns a classification slightly faster, our motion and appearance-based feature sets are able to return a single classification in less than $10ms$ and $20ms$ when the feature sets are combined, thus making it more than suitable for real-time application.

We have thus far presented separate experimental results for motion and colour features, but our ongoing work seeks to combine these feature sets to provide more robust automated species classification capable of deployment in the field to support ecological studies or migration and other population-level behaviours. This work includes not only extensions to encompass other species, but also investigations of feature selection and redundancy.

References

Anjum, N., Cavallaro, A., 2008. Multifeature object trajectory clustering for video analysis. Circuits and Systems for Video Technology, IEEE Transactions on 18, 1555–1564.

Atanbori, J., Cowling, P., Murray, J., Colston, B., Eady, P., Hughes, D., Nixon, I., Dickinson, P., 2013. Analysis of bat wing beat frequency using fourier transform, in: Computer Analysis of Images and Patterns, Springer. pp. 370–377.

Bardeli, R., Wolff, D., Kurth, F., Koch, M., Tauchert, K.H., Frommolt, K.H., 2010. Detecting bird sounds in a complex acoustic environment and application to bioacoustic monitoring. Pattern Recognition Letters 31, 1524–1534.

Bashir, F.I., Khokhar, A.A., Schonfeld, D., 2006. View-invariant motion trajectory-based activity classification and recognition. Multimedia Systems 12, 45–54.

Berg, T., Belhumeur, P.N., 2013. Poof: Part-based one-vs.-one features for fine-grained categorization, face verification, and attribute estimation, in: Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on, IEEE. pp. 955–962.

Berg, T., Liu, J., Lee, S.W., Alexander, M.L., Jacobs, D.W., Belhumeur, P.N., 2014. Birdsnap: Large-scale fine-grained visual categorization of birds, in: Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on, IEEE. pp. 2019–2026.

Beyan, C., Fisher, R.B., 2012. A filtering mechanism for normal fish trajectories, in: Pattern Recognition (ICPR), 2012 21st International Conference On, IEEE. pp. 2286–2289.

Beyan, C., Fisher, R.B., 2013. Detection of abnormal fish trajectories using a clustering based hierarchical classifier. BMVC, Bristol, UK .

Branson, S., Van Horn, G., Belongie, S., Perona, P., 2014. Bird species categorization using pose normalized deep convolutional nets. arXiv preprint arXiv:1406.2952 .

Briggs, F., Lakshminarayanan, B., Neal, L., Fern, X.Z., Raich, R., Hadley, S.J., Hadley, A.S., Betts, M.G., 2012. Acoustic classification of multiple simultaneous bird species: A multi-instance multi-label approach. The Journal of the Acoustical Society of America 131, 4640–4650.

Briggs, F., Raich, R., Fern, X.Z., 2009. Audio classification of bird species: A statistical manifold approach, in: Data Mining, 2009. ICDM'09. Ninth IEEE International Conference on, IEEE. pp. 51–60.

Buckland, S., Baillie, S., Dick, J.M., Elston, D., Magurran, A., Scott, E., Smith, R., Somerfield, P., Studeny, A., Watt, A., 2012. How should regional biodiversity be monitored? Environmental and Ecological Statistics 19, 601–626.

Claesen, M., De Smet, F., Suykens, J.A., De Moor, B., 2014. Fast prediction with svm models containing rbf kernels. arXiv preprint arXiv:1403.0736 .

Du, J.X., Wang, X.F., Zhang, G.J., 2007. Leaf shape based plant species recognition. Applied mathematics and computation 185, 883–893.

Duan, K., Parikh, D., Crandall, D., Grauman, K., 2012. Discovering localized attributes for fine-grained recognition, in: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, IEEE. pp. 3474–3481.

Duberstein, C., Virden, D., Matzner, S., Myers, J., Cullinan, V., Maxwell, A., 2012. Automated thermal image processing for detection and classification of birds and bats .

Gavves, E., Fernando, B., Snoek, C.G., Smeulders, A.W., Tuytelaars, T., 2013. Fine-grained categorization by alignments, in: Computer Vision (ICCV), 2013 IEEE International Conference on, IEEE. pp. 1713–1720.

Goodenough, A.E., Fairhurst, S.M., Morrison, J.B., Cade, M., Morgan, P.J., Wood, M.J., 2014. Quantifying the robustness of first arrival dates as a measure of avian migratory phenology. Ibis .

Gregory, R., 2006. Birds as biodiversity indicators for europe. Significance 3, 106–110.

Hammers, M., Müskens, G.J., van Kats, R.J., Teunissen, W.A., Kleijn, D., 2014. Ecological contrasts drive responses of wintering farmland birds to conservation management. Ecography .

Harrison, P.J., Buckland, S.T., Yuan, Y., Elston, D.A., Brewer, M.J., Johnston, A., Pearce-Higgins, J.W., 2014. Assessing trends in biodiversity over space and time using the example of british breeding birds. Journal of Applied Ecology 51, 1650–1660.

Hsu, C.W., Chang, C.C., Lin, C.J., et al., 2003. A practical guide to support vector classification.

Huang, C., Luo, B., Tang, L., Liu, Y., Ma, J., 2013. Topic model based bird breed classification and annotation, in: Communications, Circuits and Systems (ICCCAS), 2013 International Conference on, IEEE. pp. 319–322.

Huang, Z.C., Chan, P.P., Ng, W.W., Yeung, D.S., 2010. Content-based image retrieval using color moment and gabor texture feature, in: Machine Learning and Cybernetics (ICMLC), 2010 International Conference on, IEEE. pp. 719–724.

Johnston, A., Thaxter, C.B., Austin, G.E., Cook, A.S., Humphreys, E.M., Still, D.A., Mackay, A., Irvine, R., Webb, A., Burton, N.H., 2014. Modelling the abundance and distribution of marine birds accounting for uncertain species identification. Journal of Applied Ecology .

Keerthi, S.S., Chapelle, O., DeCoste, D., 2006. Building support vector machines with reduced classifier complexity. The Journal of Machine Learning Research 7, 1493–1515.

Lazarevic, L., Harrison, D., Southee, D., Wade, M., Osmond, J., 2008. Wind farm and fauna interaction: detecting bird and bat wing beats through cyclic motion analysis. International Journal of Sustainable Engineering 1, 60–68.

Li, X., Hu, W., Hu, W., 2006. A coarse-to-fine strategy for vehicle motion trajectory clustering, in: Pattern Recognition, 2006. ICPR 2006. 18th International Conference on, IEEE. pp. 591–594.

Liwicki, M., Bunke, H., et al., 2006. Hmm-based on-line recognition of handwritten whiteboard notes, in: Tenth International Workshop on Frontiers in Handwriting Recognition.

Lopes, M.T., Lameiras Koerich, A., Nascimento Silla, C., Alves Kaestner, C., 2011. Feature set comparison for automatic bird species identification, in: Systems, Man, and Cybernetics (SMC), 2011 IEEE International Conference on, IEEE. pp. 965–970.

Mai, F., Chang, C., Hung, Y., 2010. Affine-invariant shape matching and recognition under partial occlusion, in: Image Processing (ICIP), 2010 17th IEEE International Conference on, IEEE. pp. 4605–4608.

Marini, A., Facon, J., Koerich, A.L., 2013. Bird species classification based on color features, in: Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on, IEEE. pp. 4336–4341.

Martín, J.A., Santos, M., de Lope, J., 2010. Orthogonal variant moments features in image analysis. Information Sciences 180, 846–860.

Neal, L., Briggs, F., Raich, R., Fern, X.Z., 2011. Time-frequency segmentation of bird song in noisy acoustic environments, in: Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on, IEEE. pp. 2012–2015.

Ou, J., Bai, X.B., Pei, Y., Ma, L., Liu, W., 2010. Automatic facial expression recognition using gabor filter and expression analysis, in: Computer Modeling and Simulation, 2010. ICCMS'10. Second International Conference on, IEEE. pp. 215–218.

Parvin, H., Mohammadi, M., Rezaei, Z., 2012. Face identification based on gabor-wavelet features. JDCTA: International Journal of Digital Content Technology and its Applications 6, 247–255.

Pun, C.M., Lee, M.C., 2003. Log-polar wavelet energy signatures for rotation and scale invariant texture classification. Pattern Analysis and Machine Intelligence, IEEE Transactions on 25, 590–603.

Sergyan, S., 2008. Color histogram features based image classification in content-based image retrieval systems, in: Applied Machine Intelligence and Informatics, 2008. SAMI 2008. 6th International Symposium on, IEEE. pp. 221–224.

Spampinato, C., Giordano, D., Di Salvo, R., Chen-Burger, Y.H.J., Fisher, R.B., Nadarajan, G., 2010. Automatic fish classification for underwater species behavior understanding, in: Proceedings of the first ACM international workshop on Analysis and retrieval of tracked events and motion in imagery streams, ACM. pp. 45–50.

Suzuki, S., et al., 1985. Topological structural analysis of digitized binary images by border following. Computer Vision, Graphics, and Image Processing 30, 32–46.

Van De Sande, K.E., Gevers, T., Snoek, C.G., 2010. Evaluating color descriptors for object and scene recognition. Pattern Analysis and Machine Intelligence, IEEE Transactions on 32, 1582–1596.

Wah, C., Branson, S., Perona, P., Belongie, S., 2011a. Multiclass recognition and part localization with humans in the loop, in: Computer Vision (ICCV), 2011 IEEE International Conference on, IEEE. pp. 2524–2531.

Wah, C., Branson, S., Welinder, P., Perona, P., Belongie, S., 2011b. The caltech-ucsd birds-200-2011 dataset .

Welinder, P., Branson, S., Mita, T., Wah, C., Schroff, F., Belongie, S., Perona, P., 2010. Caltech-ucsd birds 200 .

Yao, B., Bradski, G., Fei-Fei, L., 2012. A codebook-free and annotation-free approach for fine-grained image categorization, in: Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on, IEEE. pp. 3466–3473.

Zheng, F., Webb, G.I., 2005. A comparative study of semi-naive bayes methods in classification learning, in: Proceedings of the fourth Australasian data mining conference (AusDM05), Citeseer. pp. 141–156.

Zivkovic, Z., van der Heijden, F., 2006. Efficient adaptive density estimation per image pixel for the task of background subtraction. Pattern recognition letters 27, 773–780.